

On Marine Mammal Acoustic Detection Performance Bounds

Yin Xian, Loren Nolte, Stacy Tantum, Xuejun Liao and Yuan Zhang

Abstract—Since the spectrogram does not preserve phase information contained in the original data, any algorithm based on the spectrogram is not likely to be optimum for detection. In this paper, we present the Short Time Fourier Transform detector to detect marine mammals in the time-frequency plane. The detector uses phase information for detection. We evaluate this detector by comparing it to the existing spectrogram based detectors for different SNRs and various environments including a known ocean, uncertain ocean, and mean ocean. The results show that this detector outperforms the spectrogram based detector. Simulations are presented using the polynomial phase signal model of the North Atlantic Right Whale (NARW), along with the bellhop ray tracing model.

Index Terms—Acoustic Detection, ROC, Short Time Fourier Transform, Multipath, Whale Vocalizations

I. INTRODUCTION

Many marine mammal vocalizations are non-periodic and frequency modulated signals [1], [2]. For this reason, the time-frequency analysis techniques are widely applied in analyzing such signals. Many time-frequency representation methods, such as the Short Time Fourier Transform (STFT), the Wigner Ville distribution and the wavelet transform have been used in this field. When we use the STFT, we can generate the spectrogram to examine the energy distribution of the signal in the time-frequency plane. Spectrograms are among the simplest and natural tools to analyze the AM-FM type signals [3]. It can sparsely represent the signals' energy ribbons localized along the time-frequency trajectories [4]. Many detection methods have been proposed to detect marine mammals based on the spectrogram, such as spectrogram correlation [5], [6], frequency contour edge detection method [7] and contour extraction [8], [9], [10], [11]. In order to obtain the optimal detection performance based on the spectrogram, it is natural to examine the probability distribution of the spectrogram data, and obtain the likelihood ratio for detection. Analysis of the probability distribution of the spectrogram elements has been performed to approximate the likelihood ratio of the spectrogram by assuming that the spectrogram elements are statistically independent [12]. Research can be found to perform detection based on the likelihood ratio of a single spectrogram element [13], [14].

However, the phase information is neglected when we do detection based on the spectrogram. The phase information

is particularly useful for signal reconstruction and source localization [15]. Studies have found that the phase spectrum is more sensitive than the magnitude spectrum for signal recognition when the window function of the STFT is appropriately chosen [16], [17]. By incorporating the phase information and the magnitude information, we can improve the detection performance. In this paper, we propose the STFT detector. Instead of extracting the phase information from the source signal separately for detection, we incorporate the magnitude and phase information during the STFT. Results show that the STFT detector outperforms detectors based on the spectrogram, and can achieve optimal detection result when the environmental parameters and source signal are known exactly, and the noise is additive white Gaussian. Technical details of the detector can be found in Section III and the Appendix.

The sound speed in the ocean is influenced by temperature, salinity, and pressure of the sea water, so it varies spatially [18], [19]. The speed of sound changes with depth, yielding what is known as a sound speed profile. The spatial variability gives rise to refraction of the sound. Refraction and reflection from the sea surface and bottom contribute to multipath propagation [19]. Multipath ocean propagation environment will lead to dispersion and time spread for the transmitted signal [20]. It is a challenge for the underwater acoustic communication systems. Research has been done to evaluate the uncertainty of environmental parameters to source localization [21], [22], [23], and detection in the time domain [24], [25]. In this paper, we will evaluate the influence of the uncertainty of the sound speed profile on detection in the time-frequency plane. The ROC of the time domain matched filter assuming the signal is known exactly is used to benchmark the detection performance of our proposed detector.

The organization of the paper is as follows. Section I gives an introduction and background for this paper. Section II presents the ocean acoustics propagation model, environmental parameters setting and propagated signal model. Overviews of the characteristics of the current detection model, and statistics of our proposed STFT detector are presented in Section III. Section IV gives detection performance results using the STFT detector in the matched ocean, uncertain ocean and mean ocean cases. Section V concludes the paper.

II. OCEAN ACOUSTICS PROPAGATION MODEL

A. Environment Configuration

We use the Hudson Canyon as the ocean propagation environment. The Hudson Canyon typical sound speed profile

Yin Xian, Loren Nolte, Stacy Tantum and Xuejun Liao are with the Department of Electrical and Computer Engineering, Duke University, Durham, NC, 27708 USA e-mail: (yx15@duke.edu, lwn@ee.duke.edu).

Yuan Zhang is with Department of Mathematics, UCLA, Los Angeles, CA 90095, USA email:(yuanzhang@math.ucla.edu).

is shown in Figure 1. The Hudson Canyon environment is characterized by a sandy bottom and a sound speed profile. The environmental parameters and uncertainties of the canyon are shown in Table I [21].

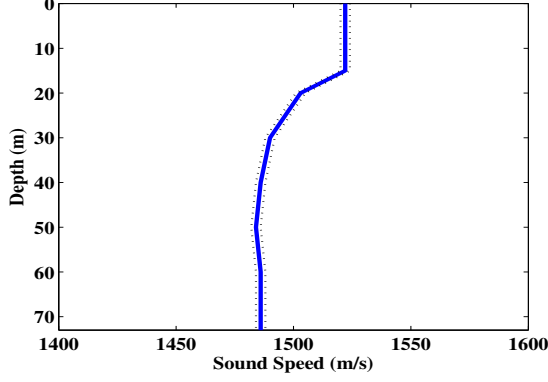


Figure 1. Hudson Canyon Typical Sound Speed Profile

Table I
HUDSON CANYON ENVIRONMENT PARAMETERS AND UNCERTAINTIES

Description	Depth	Basis sound speed	Uniform uncertainty
Water column sound speed	0+ m	1522 m/s	± 2 m/s
	15 m	1522 m/s	± 2 m/s
	20 m	1503 m/s	± 2 m/s
	30 m	1490 m/s	± 2 m/s
	40 m	1486 m/s	± 2 m/s
	50 m	1484 m/s	± 2 m/s
	60 m	1486 m/s	± 2 m/s
Bottom half-space sound speed	73- m	1486 m/s	± 2 m/s
	73+ m	1550 m/s	± 50 m/s

B. Signal Model

Many marine mammal vocalizations are frequency modulated, and can be modelled as polynomial-phase signals [2]. Take the North Atlantic Right Whale (NARW) as an example. It is found that there are nine types of sound for the NARW according to its time-frequency characteristics [26], and the upsweep call is commonly found when the whales greet each other.

Assume that $s(t)$ is the source signal of the whale, $\theta(t)$ is the phase of the signal, a is an unknown positive scalar representing the amplitude, so [2]

$$s(t) = a \cos(\theta(t)) = a \cos\left(2\pi \sum_{m=0}^{M-1} f_m t^m\right), \quad t = 1, 2, \dots, N, \quad (1)$$

where f_0, f_1, \dots, f_{M-1} are polynomial coefficient parameters for the signal model.

By inspecting the shape of the whales' sound contour, and fitting the polynomial coefficients, we can synthesize the right whale signal (Figure 2). We will use the synthetic NARW signal to evaluate the effect of propagation environment on acoustic detection performance.

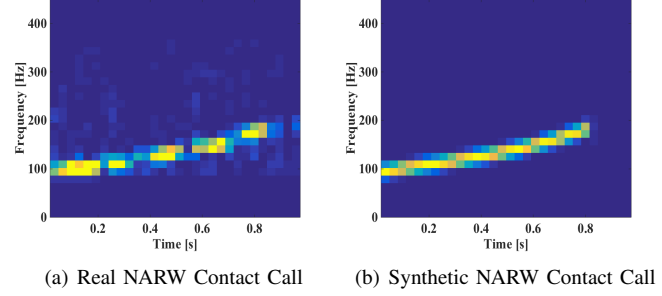


Figure 2. North Atlantic Right Whale Signal

C. Signal Propagation

Supposed that the whale is z_s meters below the sea surface. We want to determine whether there is a signal present at the receiver, which is r meters from the source and z_r meters below the sea surface. The environment setting can be characterized as in Figure 3.

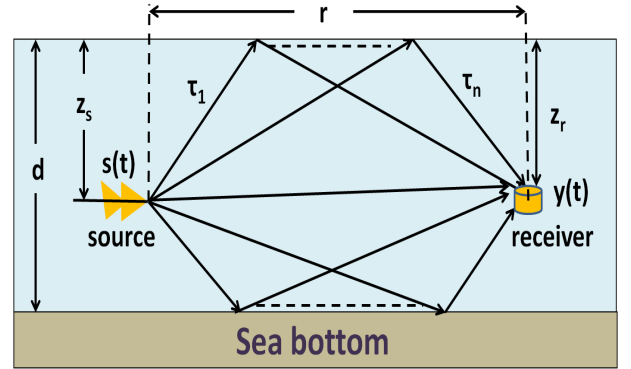


Figure 3. Ocean Acoustic Propagation Model

where,

z_s : source depth;

$s(t)$: source signal;

r : range of propagation;

τ_1 : delay of the first path;

z_r : receiver depth;

$y(t)$: propagated signal;

d : depth of shallow water;

τ_n : delay of the n-th path.

We use the Bellhop Model [27] to calculate the arrival amplitudes and delays of all possible signal paths in the ocean. Supposing there are K arrivals, the propagated signal $y(t)$ can be represented as the sum of K arrival signals. The amplitude and delay of each arrival signal is A_k and τ_k respectively, that is

$$y(t) = \sum_{k=1}^K A_k s(t - \tau_k) \quad (2)$$

where $s(t)$ is the source signal. When the amplitudes are complex numbers, which is due, for example, to bottom reflections that introduce a phase shift, the propagated signal is [28],

$$y(t) = \sum_{k=1}^K \left[\text{Re}[A_k] s(t - \tau_k) - \text{Im}[A_k] s^+(t - \tau_k) \right] \quad (3)$$

where $s^+ = \mathcal{H}(s)$ is the Hilbert transform of $s(t)$. The Hilbert transform is a 90 degrees phase shift of $s(t)$ and accounts for the imaginary part of A_k . In this case,

$$s(t) = a \cos\left(2\pi \sum_{m=0}^{M-1} f_m t^m\right) \quad (4)$$

$$s^+(t) = a \sin\left(2\pi \sum_{m=0}^{M-1} f_m t^m\right). \quad (5)$$

III. DETECTION MODELS AND STATISTICS

It is known that when the noise is additive white Gaussian, the time domain matched filter can achieve optimal detection when the source signal and the propagation environment are known exactly [19]. When examining the time-frequency performance of a given signal, we can apply the Short Time Fourier Transform to the signal, and generate the spectrogram to analyze it.

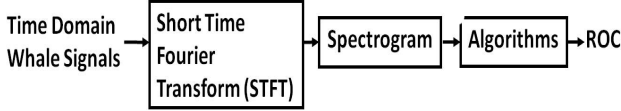


Figure 4. Detection in time-frequency domain

In this section, we will apply the likelihood ratio of the approximated probability distribution of the spectrogram [12] and present our STFT detector in the matched ocean, uncertain ocean and mean ocean cases. Simulations of comparison of detection performance can be found in Section IV.

A. Spectrogram Distribution Detector

1) *Matched ocean*: We first examine the matched ocean case, that is, the source signal and sound speed profile is known exactly. Our analysis for the matched ocean case is similar to the work of Altes [12].

Let $x(t)$, $t = 0, 1, \dots, N-1$ be the data recorded at the receiver. We want to know whether the data contains the propagated NARW signal. We can form the binary hypothesis test as follows [29]

$$H_0 : x(t) = n(t), \quad (6)$$

$$H_1 : x(t) = n(t) + y(t), \quad (7)$$

where $n(t)$ is Gaussian additive noise, $n(t) \sim \mathcal{N}(0, \sigma_n^2)$, and where $y(t)$ is the propagated NARW signal. We apply the STFT to the data and generate the spectrogram. That is [3]

$$S_x[a, b] = \left| \sum_{i=0}^{M-1} x[aD + i] w[i] \exp(-2\pi j b i / M) \right|^2, \quad (8)$$

where $w[i]$ is the window function, M is the length of the window, D is the step of the sliding window, and a is the window shift, b is the frequency shift, and where $[k]$ denotes the greatest integer less than or equal to k .

The binary hypothesis test based on the spectrogram is

$$H_0 : \mathbf{S}_x = \mathbf{S}_n, \quad (9)$$

$$H_1 : \mathbf{S}_x = \mathbf{S}_{n+y}, \quad (10)$$

where \mathbf{S}_x is the spectrogram vector of the data at the receiver, \mathbf{S}_n is the spectrogram vector of pure noise, and where \mathbf{S}_{n+y} is the spectrogram vector of the propagated signal plus noise. Let $J = \lfloor (N-M)/D \rfloor + 1$, and $B = \lfloor M/2 \rfloor + 1$, the spectrogram vector at the receiver is $\mathbf{S}_x = [S_x[0, 0], \dots, S_x[J, B]]^T$. For each spectrogram element $S_x[a, b]$ in \mathbf{S}_x , under the H_0 hypothesis, we have [30], [12], [13]

$$p(S_x[a, b]) = \frac{1}{M\sigma_n^2} \exp\left(-\frac{S_x[a, b]}{M\sigma_n^2}\right).$$

Under the H_1 hypothesis, we have [12], [13], [30]

$$p(S_x[a, b]) = \frac{1}{M\sigma_n^2} \exp\left(-\frac{S_x[a, b] + m_1^2[a, b] + m_2^2[a, b]}{M\sigma_n^2}\right) \times I_0\left(\frac{2\sqrt{(m_1^2[a, b] + m_2^2[a, b])S_x[a, b]}}{M\sigma_n^2}\right),$$

where

$$m_1[a, b] = \sum_{i=0}^{M-1} y[aD + i] \cos(2\pi b i / M),$$

$$m_2[a, b] = - \sum_{i=0}^{M-1} y[aD + i] \sin(2\pi b i / M),$$

and where $I_\nu(z)$ is the modified Bessel function of the first kind. Therefore, for each spectrogram element, the likelihood ratio is [12], [13], [30]

$$\lambda[a, b] = \frac{p(S_x[a, b]|H_1)}{p(S_x[a, b]|H_0)} = \exp\left(-\frac{m_1^2[a, b] + m_2^2[a, b]}{M\sigma_n^2}\right) \times I_0\left(\frac{2\sqrt{(m_1^2[a, b] + m_2^2[a, b])S_x[a, b]}}{M\sigma_n^2}\right). \quad (11)$$

Assuming that each element in the spectrogram matrix is statistically independent, the likelihood ratio is

$$\lambda = \frac{p(\mathbf{S}_x|H_1)}{p(\mathbf{S}_x|H_0)} = \frac{p(S_x[0, 0]|H_1) \cdots p(S_x[J, B]|H_1)}{p(S_x[0, 0]|H_0) \cdots p(S_x[J, B]|H_0)} = \prod_{a=0}^J \prod_{b=0}^B \frac{p(S_x[a, b]|H_1)}{p(S_x[a, b]|H_0)} = \prod_{a=0}^J \prod_{b=0}^B \lambda[a, b]. \quad (12)$$

2) *Uncertain ocean*: In reality, we may not know the environment and sound speed profile of the ocean exactly. If we know the prior distribution of the sound speed profile, we can apply the Bayes rule, and incorporate the prior information into detection. Suppose the sound speed profile is uniformly

distributed over P possible cases, the likelihood ratio, based on eq. (12), in this case is

$$\begin{aligned}\lambda_u &= \frac{1}{P} \sum_{k=1}^P \left(\prod_{a=0}^J \prod_{b=0}^B \lambda_k[a, b] \right) \\ &= \frac{1}{P} \sum_{k=1}^P \left(\prod_{a=0}^J \prod_{b=0}^B \left(\exp \left(-\frac{m_{k1}^2[a, b] + m_{k2}^2[a, b]}{M\sigma_n^2} \right) \right. \right. \\ &\quad \times \left. \left. I_0 \left(\frac{2\sqrt{(m_{k1}^2[a, b] + m_{k2}^2[a, b])S_{kx}[a, b]}}{M\sigma_n^2} \right) \right) \right), \quad (13)\end{aligned}$$

where λ_k is the likelihood ratio of the spectrogram of the k^{th} possible propagated signal, $S_{kx}[a, b]$ is the spectrogram element of the k^{th} possible propagated signal, and where $m_{k1}[a, b]$ and $m_{k2}[a, b]$ are the mean of real and imaginary part respectively of the k^{th} possible propagated signal.

3) *Mean ocean*: When the sound speed profile is uncertain, and we only have the knowledge of the mean value of the possible sound speed profiles, which is mismatched with the true sound speed profile. Based on eq. (12), the likelihood ratio in this case becomes [22]

$$\begin{aligned}\lambda_m &= \prod_{a=0}^J \prod_{b=0}^B \left(\exp \left(-\frac{\hat{m}_1^2[a, b] + \hat{m}_2^2[a, b]}{M\sigma_n^2} \right) \right. \\ &\quad \times \left. I_0 \left(\frac{2\sqrt{(\hat{m}_1^2[a, b] + \hat{m}_2^2[a, b])S_x[a, b]}}{M\sigma_n^2} \right) \right), \quad (14)\end{aligned}$$

where $S_x[a, b]$ is the spectrogram element of the received signal, and $\hat{m}_1[a, b]$ and $\hat{m}_2[a, b]$ are the mean value of the real part and imaginary part respectively of the Short Time Fourier Transform of the signal in the mean ocean environment, and where the sound speed profile is assumed to take average values.

B. STFT Detector

The detection algorithms based on spectrogram data neglect phase information, and therefore their detection performances are not likely to be optimal. In this section we derive the probability distribution and likelihood ratio for detection of STFT. The STFT detector preserves the phase information of the detected signal, and achieves optimal detection performance in the matched ocean case.

1) *Matched ocean*: The binary hypothesis test based on the STFT in this case is [29]

$$H_0 : X = X_n, \quad (15)$$

$$H_1 : X = X_y + X_n, \quad (16)$$

where X is the vectorized STFT matrix of the received data, X_n is the vectorized STFT matrix of pure noise, and X_y is the vectorized STFT matrix of the propagated signal. $X = [X[0, 0], \dots, X[J, B]]^T$. For each STFT element $X[a, b]$ in X , we have [3]

$$X[a, b] = \sum_{i=0}^{M-1} x[aD + i]w[i] \exp(-2\pi jbi/M), \quad (17)$$

where $a = 0, \dots, J$ and $b = 0, \dots, B$. Since $X[a, b]$ is a complex value, letting

$$U[a, b] = \text{Re}\{X[a, b]\}, \quad (18)$$

$$V[a, b] = \text{Im}\{X[a, b]\}, \quad (19)$$

we have $X[a, b] = U[a, b] + jV[a, b]$.

Concatenating the elements in the STFT matrix of the received data, and separating the real and imaginary part of each element, we can obtain

$$\mathbf{X} = (U[0, 0], \dots, U[J, B], V[0, 0], \dots, V[J, B])^T. \quad (20)$$

Because we use the Gaussian distribution as our noise model, \mathbf{X} follows a multivariate normal distribution under both H_0 and H_1 hypotheses. It can be proved that the signal covariance matrix \mathbf{C}_x is the same under both H_0 and H_1 hypotheses. The expression for \mathbf{C}_x is derived in Appendix A.

When \mathbf{C}_x is singular, the probability density function for \mathbf{X} does not exist. According to the derivation in Appendix A, mapping \mathbf{X} into a subspace formed by Q_1 , where Q_1 is the eigenvector of the non-zero singular values' component, we can obtain the probability density function for $Q_1^T \mathbf{X}$ under H_0 and H_1 hypotheses

$$\begin{aligned}p(Q_1^T \mathbf{X} | H_0) &= \frac{1}{(2\pi)^{(B+1)(J+1)} \det^{1/2}(\Lambda_1)} \\ &\quad \times \exp \left(-\frac{1}{2} \mathbf{X}^T Q_1 \Lambda_1^{-1} Q_1^T \mathbf{X} \right), \quad (21)\end{aligned}$$

$$\begin{aligned}p(Q_1^T \mathbf{X} | H_1) &= \frac{1}{(2\pi)^{(B+1)(J+1)} \det^{1/2}(\Lambda_1)} \\ &\quad \times \exp \left(-\frac{1}{2} (\mathbf{X} - \mu)^T Q_1 \Lambda_1^{-1} Q_1^T (\mathbf{X} - \mu) \right). \quad (22)\end{aligned}$$

where Λ_1 is the diagonal matrix of eigenvalues corresponding to Q_1 , and μ is the expected value of \mathbf{X} .

The likelihood ratio based on the STFT is then

$$\begin{aligned}\lambda &= \frac{p(Q_1^T \mathbf{X} | H_1)}{p(Q_1^T \mathbf{X} | H_0)} \\ &= \exp \left(-\frac{1}{2} (\mu^T Q_1 \Lambda_1^{-1} Q_1^T \mu - 2\mathbf{X}^T Q_1 \Lambda_1^{-1} Q_1^T \mu) \right). \quad (23)\end{aligned}$$

We can calculate the analytic solution for the ROC plot based on eq. (23). According to the derivation in Appendix A, the square of separation parameter, or the detection index is

$$d^2 = \mu^T Q_1 \Lambda_1^{-1} Q_1^T \mu. \quad (24)$$

2) *Uncertain ocean*: When the sound speed profile is uncertain, and is uniformly distributed over P possible cases. Applying the Bays rule, the likelihood ratio can be calculated as

$$\lambda_u = \frac{1}{P} \sum_{k=1}^P \exp \left(\mathbf{X}^T Q_1 \Lambda_1^{-1} Q_1^T \mu_k - \frac{1}{2} \mu_k^T Q_1 \Lambda_1^{-1} Q_1^T \mu_k \right), \quad (25)$$

where μ_k is the expected value of the k^{th} possible propagated signal's STFT vector.

3) *Mean ocean*: When the sound speed profile is uncertain, and the prior information of the sound speed profile is unknown, but we know the mean sound speed profile, we can cross-correlate the possible propagated signals with the signal corresponding to the mean sound speed profile environment. The likelihood ratio can be computed as [22]

$$\lambda_m = \exp\left(-\frac{1}{2}\left(\mu_m^T Q_1 \Lambda_1^{-1} Q_1^T \mu_m - 2\mathbf{X}^T Q_1 \Lambda_1^{-1} Q_1^T \mu_m\right)\right), \quad (26)$$

where μ_m is the expected value of the mean sound speed profile with respect to \mathbf{X} . We can mathematically prove that when the uncertainty of the sound speed profile is small, given that the mapping of sound speed profile to the propagated signal is linear, according to eq. (3), eq. (26) is actually the geometric mean of the likelihood ratio. Detailed proof can be found in Appendix A.

IV. RESULTS

Based on recordings of NARW from the years 2001 to 2003 in the Cape Cod Bay region, using bottom-mounted hydrophones with 2000Hz sampling rate, the polynomial coefficient set (f_0, f_1, f_2) was found most frequently to take the value of (100,0,48) [2]. Because the NARW upsweep call usually lasts for about 1 second, we let the duration of the signal be 1.024s. Substituting these values into eq. (1), and letting $a = 1$, the source signal's waveform and spectrogram can be determined and are illustrated in Figure 5.

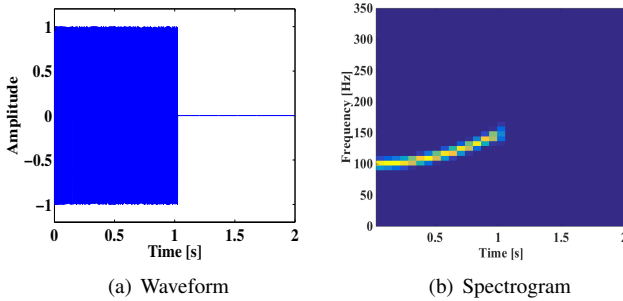


Figure 5. Synthetic NARW source signal

We place the source signal 15 meters below the sea surface, and place the receiver 35 meters below. The horizontal distance between the receiver and source signal is 1000 meters. Considering the uncertainty of the sound speed profile, we suppose that the sound speed profile has 50 possible cases. When the sound speed profile changes, the propagated signal will also change. Figure 6 illustrates an example that a source under two distinct sound speed profiles will produce two different propagated signals.

According to the spectrogram plots, we can see that there is a dispersion effect during signal propagation. Some frequency components in the source signal have been lost due to phase cancellation effects of multipath propagation environment. However, the overall shapes of 50 possible propagated signal spectrograms are similar to the source signal spectrogram; this is because the time shift will not change the frequency

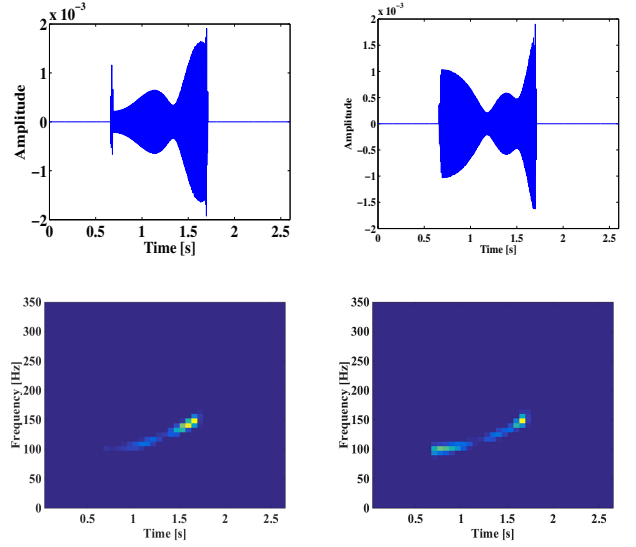


Figure 6. Example plots of time domain waveform and spectrogram of NARW propagated signal under two different sound speed profiles.

content of the signal, due to the basic property of the Fourier Transform.

A. Detectors Tested

We compare the detection performance of the STFT detector, the spectrogram distribution detector, and the spectrogram correlation detector in the matched ocean, uncertain ocean and mean ocean cases. We benchmark the results with the optimal detection performance given by the time domain matched filtering assuming that the source signal and propagation environment are known exactly.

The spectrogram correlation detector is a well known work in marine mammal acoustic detection field. Review of the spectrogram correlation detector [5] proposed by Mellinger and Clark is as follows. The spectrogram correlation constructs a kernel function for the vocalization signal, then cross correlates it with the target signal's spectrogram to calculate the recognition score, and does detection based on the recognition score. The kernel function for the signal is made up of several segments, one per FM section in the target vocalization type. The kernel value k at a given time and frequency point (t, f) is specified by:

$$x = f - \left(f_0 + \frac{t}{d}(f_1 - f_0)\right)$$

$$k(t, f) = \left(1 - \frac{x^2}{\sigma^2}\right) \exp\left(-\frac{x^2}{2\sigma^2}\right)$$

where x is the distance of the point (t, f) from the central axis of the segment at time t , f_0 is the start frequency of the segment, f_1 is the end frequency of the segment, d is the duration of the segment, and σ is the instantaneous bandwidth of the segment at time t . The recognition score is calculated by cross correlating the kernel $k(t, f)$ with the spectrogram of the signal:

$$\alpha(t) = \sum_{t_0} \sum_f k(t_0, f) S(t - t_0, f)$$

We use a rectangular window function with length 256, and overlap size 128 to generate the spectrogram of the signal. We apply these parameters to the spectrogram correlation detector and spectrogram distribution detector. We set a threshold for the recognition score and generate the Receiver Operating Characteristic (ROC) curve.

The comparison of detection performance of the STFT detector, the spectrogram distribution detector and the spectrogram correlation detector is shown in Figure 7. We use a Monte Carlo method to do the numerical simulation. We will analyze the detection performance in the matched ocean, uncertain ocean and mean ocean cases.

B. Matched Ocean

We first examine the special case of detection in a matched ocean propagation environment. We can see that in this case the STFT detector achieves identical detection performance to the time domain matched filter case, which is the optimal detection performance. We used different window lengths and amounts of overlap, and the STFT detector performs identically in all situations. Figure 8 shows the analytic ROC plots in different SNRs for the STFT detector, based on the separation index in eq. (24) and the analysis in Appendix A; along with the corresponding analytic time domain matched filtered ROC curves as a comparison. We can see the analytic derivation and the numerical simulation are a good match.

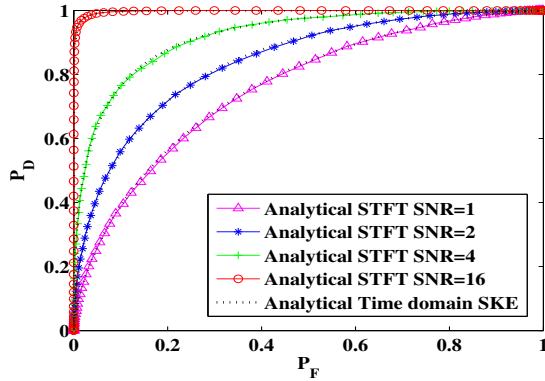


Figure 8. ROC plot based on analytical derivation when the environmental parameters are known exactly in different SNRs

Examining Figure 7, the STFT detector performs better than both the spectrogram distribution detector and the spectrogram correlation detector under different SNRs since it preserves the phase information. The spectrogram distribution detector performs better than the spectrogram correlation detector, and the spectrogram correlation detector does not perform well especially at low SNR. This is because the multipath effect will lead to dispersion for the signal, and some energy of the signal will be lost due to phase cancelation. When such a propagated signal is corrupted by noise, the performance of the spectrogram correlation detector suffers further, because it doesn't exploit the probability distribution of the spectrogram elements and noise.

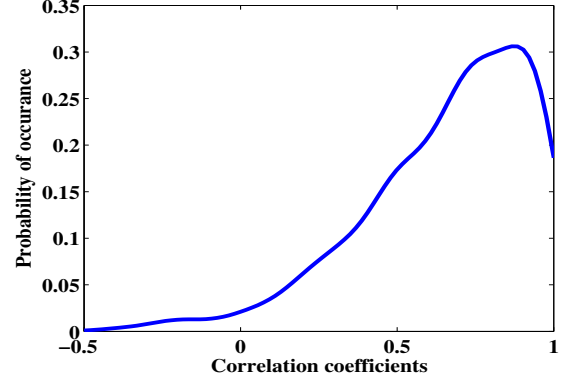


Figure 9. Kernel density function of propagated signals' correlation coefficients of different sound speed profiles

C. Uncertain Ocean

For the uncertain ocean case, we assume there are 50 possible sets of sound speed profiles. The values of the sound speed profile are given in Table I. From the ROC plots, we can see that the STFT detector performs close to the optimal case given by the time domain matched filter; and its detection performance is much better than the spectrogram distribution detector and spectrogram correlation detector. The kernel density function of correlation coefficients of the propagated signals of each possible sound speed profile is shown in Figure 9. We can see that most of the correlation coefficients are above 0.5; therefore, when we assume a uniform distribution over the possible sound speed profiles, the STFT detector performs close to the optimum.

D. Mean Ocean

For the mean ocean case, we can see that the overall detection performance deteriorates. Comparing the figures for the matched ocean case, the uncertain ocean case and the mean ocean case, we can see that the detection performance for the matched ocean case is the best, and the uncertain ocean is better than the mean ocean case. This is because the matched ocean has the most information while the mean ocean has the least. The performance of the spectrogram distribution detector does not change as much as the STFT detector, and there is little change in the performance of the spectrogram correlation detector for the three different environment cases. This implies that, by neglecting phase information, the detector will be less sensitive to environmental uncertainty.

V. CONCLUSION

In this chapter, we have presented the STFT detector, which preserves the phase information of the signal. We have also determined the likelihood ratio of the STFT detector and the spectrogram distribution detector in the matched ocean, uncertain ocean and mean ocean cases. Experiments show that the STFT detector performs better than the spectrogram distribution detector and spectrogram correlation detector. The

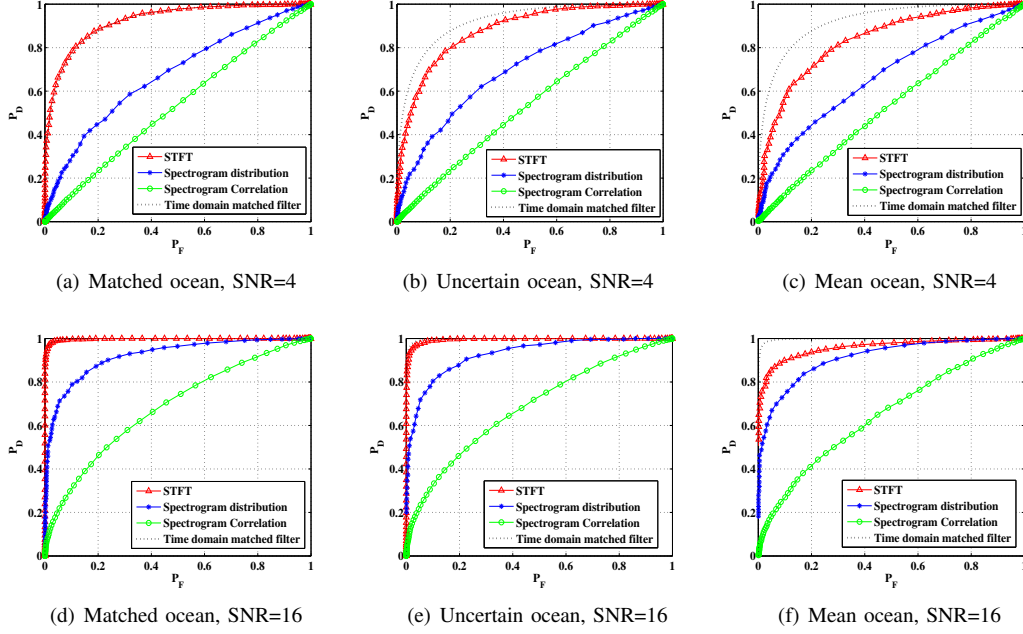


Figure 7. ROC plots for the detectors in various ocean environments and SNRs. (a), (b) and (c) show the detectors performances of the matched ocean, uncertain ocean and mean ocean environment case when SNR=4; (d), (e) and (f) show the matched ocean, uncertain ocean and mean ocean environment case when SNR=16.

STFT detector is more sensitive to environmental changes, because it includes the phase information, while the spectrogram based detectors are less sensitive. By exploiting the probability distribution of noise and spectrogram element, the detector can be improved and be more robust to multipath propagation environment.

APPENDIX A DERIVATION OF SHORT TIME FOURIER TRANSFORM DETECTOR

A. Matched ocean

According to eq. (20), we can see that \mathbf{X} is of dimension $(2(B+1) \times (J+1)) \times 1$. Because we use the Gaussian distribution as our noise model, \mathbf{X} follows a multivariate normal distribution under both H_0 and H_1 hypotheses.

Under the H_0 hypothesis, we have

$$\mathbb{E}(\mathbf{X}) = \mathbf{0}_{2(B+1) \times (J+1) \times 1}. \quad (27)$$

The covariance matrix \mathbf{C}_x of \mathbf{X} can be represented as [31]

$$\mathbf{C}_x = \begin{bmatrix} \mathbf{C}_{UU} & \mathbf{C}_{UV} \\ \mathbf{C}_{VU} & \mathbf{C}_{VV} \end{bmatrix}, \quad (28)$$

where

$$\mathbf{C}_{UU} = \begin{bmatrix} \text{Var}(U[0,0]) & \cdots & \text{Cov}(U[0,0], U[J,B]) \\ \vdots & \ddots & \vdots \\ \text{Cov}(U[J,B], U[0,0]) & \cdots & \text{Var}(U[J,B]) \end{bmatrix}, \quad \mathbf{C}_{VV} = \begin{bmatrix} \text{Var}(V[0,0]) & \cdots & \text{Cov}(V[0,0], V[J,B]) \\ \vdots & \ddots & \vdots \\ \text{Cov}(V[J,B], V[0,0]) & \cdots & \text{Var}(V[J,B]) \end{bmatrix},$$

$$\mathbf{C}_{UV} = \begin{bmatrix} \text{Cov}(U[0,0], V[0,0]) & \cdots & \text{Cov}(U[0,0], V[J,B]) \\ \vdots & \ddots & \vdots \\ \text{Cov}(U[J,B], V[0,0]) & \cdots & \text{Cov}(U[J,B], V[J,B]) \end{bmatrix},$$

$$\mathbf{C}_{VU} = \begin{bmatrix} \text{Cov}(V[0,0], U[0,0]) & \cdots & \text{Cov}(V[0,0], U[J,B]) \\ \vdots & \ddots & \vdots \\ \text{Cov}(V[J,B], U[0,0]) & \cdots & \text{Cov}(V[J,B], U[J,B]) \end{bmatrix}.$$

For $a_1, a_2 = 0, 1, \dots, J$ and $b_1, b_2 = 0, 1, \dots, B$, applying the rectangular window function $w[i] = 1$, for $i = 0, \dots, M-1$, we have

$$\begin{aligned} & \text{Cov}(U[a_1, b_1], U[a_2, b_2]) \\ &= \sigma_n^2 \sum_{i_1=0}^{M-1} \sum_{i_2=0}^{M-1} \delta[a_1 D + i_1 - a_2 D - i_2] \cos\left(\frac{2\pi b_1 i_1}{M}\right) \cos\left(\frac{2\pi b_2 i_2}{M}\right) \\ & \text{Cov}(V[a_1, b_1], V[a_2, b_2]) \\ &= \sigma_n^2 \sum_{i_1=0}^{M-1} \sum_{i_2=0}^{M-1} \delta[a_1 D + i_1 - a_2 D - i_2] \sin\left(\frac{2\pi b_1 i_1}{M}\right) \sin\left(\frac{2\pi b_2 i_2}{M}\right) \\ & \text{Cov}(U[a_1, b_1], V[a_2, b_2]) \\ &= -\sigma_n^2 \sum_{i_1=0}^{M-1} \sum_{i_2=0}^{M-1} \delta[a_1 D + i_1 - a_2 D - i_2] \cos\left(\frac{2\pi b_1 i_1}{M}\right) \sin\left(\frac{2\pi b_2 i_2}{M}\right) \end{aligned}$$

$$\begin{aligned} & \text{Cov}(V[a_1, b_1], U[a_2, b_2]) \\ &= -\sigma_n^2 \sum_{i_1=0}^{M-1} \sum_{i_2=0}^{M-1} \delta[a_1 D + i_1 - a_2 D - i_2] \sin\left(\frac{2\pi b_1 i_1}{M}\right) \cos\left(\frac{2\pi b_2 i_2}{M}\right). \end{aligned}$$

When the covariance matrix is singular, we apply spectral decomposition to the covariance matrix

$$\mathbf{C}_x = \mathbf{Q} \mathbf{\Lambda} \mathbf{Q}^T$$

where Q is a unitary orthogonal matrix, and Λ is a diagonal matrix. The diagonal elements in Λ are all real-valued. Suppose there are k non-zero eigenvalues in Λ , then the covariance matrix C_x is:

$$\begin{aligned} C_x &= Q\Lambda Q^T \\ &= [Q_1 \ Q_2] \begin{bmatrix} \Lambda_1 & 0 \\ 0 & \Lambda_2 \end{bmatrix} \begin{bmatrix} Q_1^T \\ Q_2^T \end{bmatrix} \\ &= Q_1\Lambda_1Q_1^T + Q_2\Lambda_2Q_2^T \end{aligned}$$

where Λ_1 and Q_1 corresponds to the k non-zero eigenvalues components, and Λ_2 and Q_2 corresponds to the zero components.

When C_x is singular, the probability density function for \mathbf{X} does not exist. Mapping \mathbf{X} into a subspace formed by Q_1 , the probability density function for $Q_1^T \mathbf{X}$ is

$$\begin{aligned} p(Q_1^T \mathbf{X}) &= \frac{1}{(2\pi)^{(B+1)(J+1)} \det^{1/2}(\Lambda_1)} \\ &\times \exp\left(-\frac{1}{2} \mathbf{X}^T Q_1 \Lambda_1^{-1} Q_1^T \mathbf{X}\right). \end{aligned} \quad (29)$$

Under the H_1 hypothesis, letting

$$\mu = \mathbb{E}(\mathbf{X}),$$

we know that

$$\begin{aligned} \mu &= \left[\sum_{i=0}^{M-1} y[0 \cdot D + i] \cos(2\pi 0 \cdot i/M), \dots, \right. \\ &\quad \sum_{i=0}^{M-1} y[J \cdot D + i] \cos(2\pi(M-1) \cdot i/M), \\ &\quad \left. \sum_{i=0}^{M-1} y[0 \cdot D + i] \sin(-2\pi 0 \cdot i/M), \dots, \right. \\ &\quad \left. \sum_{i=0}^{M-1} y[J \cdot D + i] \sin(-2\pi(M-1) \cdot i/M) \right]^T \end{aligned}$$

and,

$$\mathbb{E}(Q_1^T \mathbf{X}) = Q_1^T \mu.$$

It can be proved that the covariance matrix C_x under the H_1 hypothesis is the same as the one under the H_0 hypothesis. Therefore, the probability density function under the H_1 hypothesis for $Q_1^T \mathbf{X}$ is

$$\begin{aligned} p(Q_1^T \mathbf{X}) &= \frac{1}{(2\pi)^{(B+1)(J+1)} \det^{1/2}(\Lambda_1)} \\ &\times \exp\left(-\frac{1}{2} (\mathbf{X} - \mu)^T Q_1 \Lambda_1^{-1} Q_1^T (\mathbf{X} - \mu)\right). \end{aligned} \quad (30)$$

Based on eq. (29) and eq. (30), we find the likelihood ratio based on STFT data, when the environmental parameters and source signal is known exactly, to be

$$\begin{aligned} \lambda &= \frac{p(Q_1^T \mathbf{X} | H_1)}{p(Q_1^T \mathbf{X} | H_0)} \\ &= \exp\left(-\frac{1}{2} (\mu^T Q_1 \Lambda_1^{-1} Q_1^T \mu - 2 \mathbf{X}^T Q_1 \Lambda_1^{-1} Q_1^T \mu)\right). \end{aligned} \quad (31)$$

Therefore, the log likelihood ratio is

$$\ln \lambda = -\frac{1}{2} \left(\mu^T Q_1 \Lambda_1^{-1} Q_1^T \mu - 2 \mathbf{X}^T Q_1 \Lambda_1^{-1} Q_1^T \mu \right). \quad (32)$$

Because \mathbf{X} follows a multivariate normal distribution under both hypotheses, $Q_1^T \mathbf{X}$ also follows a multivariate normal distribution. We can derive the log likelihood distribution under both hypothesis, and derive the analytic solution for the ROC plots:

$$\begin{aligned} \mu_1 &= \mathbb{E}(\ln \lambda | H_1) = \frac{1}{2} (\mu^T Q_1 \Lambda_1^{-1} Q_1^T \mu) \\ \sigma_1^2 &= \text{Var}(\ln \lambda | H_1) \\ &= (Q_1 \Lambda_1^{-1} Q_1^T \mu)^T \text{Cov}(\mathbf{X}) Q_1 \Lambda_1^{-1} Q_1^T \mu \\ &= (Q_1 \Lambda_1^{-1} Q_1^T \mu)^T (Q_1 \Lambda_1 Q_1^T + Q_2 \Lambda_2 Q_2^T) Q_1 \Lambda_1^{-1} Q_1^T \mu \\ &= \mu^T Q_1 \Lambda_1^{-1} Q_1^T \mu \\ \mu_0 &= \mathbb{E}(\ln \lambda | H_0) = -\frac{1}{2} (\mu^T Q_1 \Lambda_1^{-1} Q_1^T \mu) \\ \sigma_0^2 &= \text{Var}(\ln \lambda | H_0) = \text{Var}(\ln \lambda | H_1) = \mu^T Q_1 \Lambda_1^{-1} Q_1^T \mu. \end{aligned}$$

Therefore,

$$\ln \lambda | H_1 \sim \mathcal{N}\left(\frac{1}{2} (\mu^T Q_1 \Lambda_1^{-1} Q_1^T \mu), \mu^T Q_1 \Lambda_1^{-1} Q_1^T \mu\right) \quad (33)$$

$$\ln \lambda | H_0 \sim \mathcal{N}\left(-\frac{1}{2} (\mu^T Q_1 \Lambda_1^{-1} Q_1^T \mu), \mu^T Q_1 \Lambda_1^{-1} Q_1^T \mu\right). \quad (34)$$

The probability of correct detection P_D and the probability of false alarm P_F in this case are [29]:

$$\begin{aligned} P_D &= \int_{\ln \beta}^{\infty} p(\ln \lambda | H_1) d(\ln \lambda) \\ &= \frac{1}{\sqrt{2\pi\sigma_1^2}} \int_{\ln \beta}^{\infty} e^{-\frac{(\ln \lambda - \mu_1)^2}{2\sigma_1^2}} d(\ln \lambda) \\ P_F &= \int_{\ln \beta}^{\infty} p(\ln \lambda | H_0) d(\ln \lambda) \\ &= \frac{1}{\sqrt{2\pi\sigma_0^2}} \int_{\ln \beta}^{\infty} e^{-\frac{(\ln \lambda - \mu_0)^2}{2\sigma_0^2}} d(\ln \lambda). \end{aligned}$$

Since $\sigma_1^2 = \sigma_0^2$, the standard deviation in this case is positive, so $\sigma_1 = \sigma_0$. Letting $\sigma = \sigma_1 = \sigma_0$, $z = \frac{(\ln \lambda - \mu_1)}{\sigma}$, $z' = \frac{(\ln \lambda - \mu_0)}{\sigma}$, $\sigma' = (\ln \beta - \mu_0)/\sigma$, so $dz = \sigma d(\ln \lambda)$, $dz' = \sigma d(\ln \lambda)$, we have

$$P_F = \frac{1}{\sqrt{2\pi}} \int_{\sigma'}^{\infty} e^{-\frac{z'^2}{2}} dz', \quad (35)$$

$$P_D = \frac{1}{\sqrt{2\pi}} \int_{\sigma' - \frac{\mu_1 - \mu_0}{\sigma}}^{\infty} e^{-\frac{z^2}{2}} dz, \quad (36)$$

and the separation parameter, or detection index, d in this case is

$$d = \frac{\mu_1 - \mu_0}{\sigma},$$

where,

$$\begin{aligned} d^2 &= \frac{(\mu_1 - \mu_0)^2}{(\sigma)^2} \\ &= \frac{(\mu^T Q_1 \Lambda_1^{-1} Q_1^T \mu)^2}{\mu^T Q_1 \Lambda_1^{-1} Q_1^T \mu} = \mu^T Q_1 \Lambda_1^{-1} Q_1^T \mu. \end{aligned} \quad (37)$$

B. Mean ocean

The following is a proof that when the uncertainty of the sound speed profile is small, we can theoretically approximate the likelihood ratio using eq. (26). Suppose p_k is the probability of the sound speed profile of the k^{th} path; μ_k is the propagated signal corresponding to the k^{th} sound speed profile, and μ_m is the propagated signal of the mean sound speed profile. According to eq. (3), we can see that the mapping of the sound speed and propagated signal is linear, so we have $\mu_m = \sum_{k=1}^P p_k \mu_k$. According to eq. (23), the geometric mean of the likelihood ratio is

$$\begin{aligned} & \prod_{k=1}^P \lambda_k^{p_k} \\ &= \exp \left(-\frac{1}{2} \left(\sum_{k=1}^P p_k \left(\mu_k^T Q_1 \Lambda_1^{-1} Q_1^T \mu_k - 2 \mathbf{X}^T Q_1 \Lambda_1^{-1} Q_1^T \mu_k \right) \right) \right) \\ &= \exp \left(\left(-\frac{1}{2} \sum_{k=1}^P p_k \mu_k^T Q_1 \Lambda_1^{-1} Q_1^T \mu_k \right) + \mathbf{X}^T Q_1 \Lambda_1^{-1} Q_1^T \mu_m \right) \end{aligned}$$

For the term $-\frac{1}{2} \sum_{k=1}^P p_k \mu_k^T Q_1 \Lambda_1^{-1} Q_1^T \mu_k$, notice that:

$$\begin{aligned} & \sum_{k=1}^P p_k \mu_k^T Q_1 \Lambda_1^{-1} Q_1^T \mu_k \\ &= \sum_{k=1}^P p_k \mu_k^T Q_1 \Lambda_1^{-1} Q_1^T (\mu_k - \mu_m) \\ &= \sum_{k=1}^P p_k (\mu_k^T - \mu_m^T) Q_1 \Lambda_1^{-1} Q_1^T (\mu_k - \mu_m) \\ & \quad + \sum_{k=1}^P p_k \mu_m^T Q_1 \Lambda_1^{-1} Q_1^T (\mu_k - \mu_m) \\ &= \sum_{k=1}^P p_k (\mu_k^T - \mu_m^T) Q_1 \Lambda_1^{-1} Q_1^T (\mu_k - \mu_m) \\ &\leq C \sum_{k=1}^P p_k \|\mu_k - \mu_m\|^2 = C \text{Var}(\mu_k) \end{aligned}$$

where C is the largest eigenvalue of Λ_1^{-1} . When the variance of μ_k is small, we can approximate $\sum_{k=1}^P p_k \mu_k^T Q_1 \Lambda_1^{-1} Q_1^T \mu_k$ with $\sum_{k=1}^P p_k \mu_m^T Q_1 \Lambda_1^{-1} Q_1^T \mu_m$. Therefore,

$$\prod_{k=1}^P \lambda_k^{p_k} \approx \exp \left(-\frac{1}{2} \left(\mu_m^T Q_1 \Lambda_1^{-1} Q_1^T \mu_m - 2 \mathbf{X}^T Q_1 \Lambda_1^{-1} Q_1^T \mu_m \right) \right).$$

We can see that μ_m in eq. (26) is the geometric mean of the likelihood ratio.

ACKNOWLEDGMENT

The authors would like to thank Dr. Surya Tokdar, Dr. Andrew Thompson, Dr. Granger Hickman, Dr. Yingbo Li, Dr. Galen Reeves for helpful discussion and instructions.

REFERENCES

- [1] M. D. Beecher. "Spectrographic analysis of animal vocalizations: implications of the 'uncertainty principle'," *Bioacoustics* **1**, 187-208 (1988).
- [2] I. R. Urazghildiev, and C. W. Clark. "Acoustic detection of North Atlantic right whale contact calls using the generalized likelihood ratio test," *J. Acoust. Soc. Am.* **120**, 1956-1963 (2006).
- [3] P. Flandrin. *Time-Frequency/Time-Scale Analysis* (Academic Press, San Diego, CA, 1999), pp. 49-58.
- [4] P. Flandrin, and P. Borgnat. "Time-frequency energy distributions meet compressed sensing," *IEEE Trans. Signal Process.* **58**, 2974-2982 (2010).
- [5] D. K. Mellinger, and C. W. Clark. "Recognizing transient low-frequency whale sounds by spectrogram correlation," *J. Acoust. Soc. Am.* **107**, 3518-3529 (2000).
- [6] D. K. Mellinger, and C. W. Clark. "Methods for automatic detection of mysticete sounds," *Mar. Fresh. Behav. Physiol.* **29**, 163-181 (1997).
- [7] D. Gillespie. "Detection and classification of right whale calls using an 'edge' detector operating on a smoothed spectrogram," *Can. Acoust.* **32**, 39-47 (2004).
- [8] D. K. Mellinger, S. W. Martin, R. P. Morrissey, L. Thomas, and J. J. Yosco. "A method for detecting whistles, moans, and other frequency contour sounds," *J. Acoust. Soc. Am.* **129**, 4055-4061 (2011).
- [9] J. R. Buck, and P. L. Tyack. "A quantitative measure of similarity for tursiops truncatus signature whistles," *J. Acoust. Soc. Am.* **94**, 2497-2506 (1993).
- [10] S. K. Madhusudhana, E. M. Oleson, M. S. Soldevilla, M. Roch, and J. Hildebrand. "Frequency based algorithm for robust contour extraction of blue whale B and D calls," *OCEAN 2008-MTS/IEEE Kobe Techno-Ocean*, pp. 1-8 (2008).
- [11] A. Mallawaarachchi, S. H. Ong, M. Chitre, and E. Taylor. "Spectrogram denoising and automated extraction of the fundamental frequency variation of dolphin whistles," *J. Acoust. Soc. Am.* **124**, 1159-1170 (2008).
- [12] R. A. Atlas. "Detection, estimation and classification with spectrogram," *J. Acoust. Soc. Am.* **67**, 1232-1245 (1980).
- [13] J. Huillery, F. Millioz, and N. Martin. "On the description of spectrogram probabilities with a chi-squared law," *IEEE Trans. Signal Process.* **56**, 2249-2258 (2008).
- [14] J. Huillery, F. Millioz, and N. Martin. "On the probability distributions of spectrogram coefficients for correlated Gaussian process," *Proc. IEEE Int. Conf. Acoust. Speech Signal Process.* Vol. 3, pp. 436-439 (2006).
- [15] A. V. Oppenheim, and J. S. Lim. "The importance of phase in signals," *Proc. IEEE* **69**, 529-541 (1981).
- [16] L. Liu, J. He, and G. Palm. "Effects of phase on the perception of intervocalic stop consonants," *Speech Commun.* **22**, 403-417 (1997).
- [17] K. K. Paliwal, and L. D. Alsteris. "On the usefulness of STFT phase spectrum in human listening tests," *Speech Commun.* **45**, 153-170 (2005).
- [18] J. Preisig. "Acoustic propagation considerations for underwater acoustic communications network development," *SIGMOBILE* **11**, 2-10 (2007).
- [19] D. H. Johnson, and D. E. Dudgeon. *Array signal processing: concepts and techniques* (Prentice Hall PTR, Upper Saddle River, NJ, 1993), pp. 10-45.
- [20] D. Rouseff, D. R. Jackson, W. L. Fox, C. D. Jones, J. A. Ritcey and D. R. Dowling. "Underwater acoustic communication by passive-phase conjugation: Theory and experimental results," *IEEE J. Ocean. Eng.* **26**, 821-831 (2001).
- [21] J. A. Shorey, and L. W. Nolte. "Wideband optimal a posteriori probability source localization in an uncertain shallow ocean environment," *J. Acoust. Soc. Am.* **103**, 355-361 (1998).
- [22] P. J. Book, and L. W. Nolte. "Narrow-band source localization in the presence of internal waves for 1000-km range and 25-Hz acoustic frequency," *J. Acoust. Soc. Am.* **101**, 1336-1346 (1997).
- [23] S. L. Tantum, and L. W. Nolte. "On array design for matched-field processing," *J. Acoust. Soc. Am.* **107**, 2101-2111 (2000).
- [24] L. Sha, and L. W. Nolte. "Effects of environmental uncertainties on sonar detection performance prediction," *J. Acoust. Soc. Am.* **117**, 1942-1953 (2005).
- [25] M. Wazenski, and D. Alexandrou. "Active wideband detection and localization in an uncertain multipath environment," *J. Acoust. Soc. Am.* **101**, 1961-1970 (1997).

- [26] V. Trygonis, E. Gerstein, J. Moir, and S. McCulloch. "Vocalization characteristics of North Atlantic right whale surface active groups in the calving habitat, southeastern United States," *J. Acoust. Soc. Am.* **134**, 4518-4531 (2013).
- [27] M. B. Porter, "The BELLHOP manual and user's guide: Preliminary draft," <http://oalib.hlsresearch.com/Rays/HLS-2010-1.pdf> (Last viewed 6/24/15).
- [28] M. Siderius, and M. B. Porter. "Modeling broadband ocean acoustic transmissions with time-varying sea surfaces," *J. Acoust. Soc. Am.* **124**, 137-150 (2008).
- [29] S. M. Kay. *Fundamentals of Statistical Signal Processing: Detection Theory* (Prentice Hall PTR, Upper Saddle River, NJ, 1998), pp. 8-15, 60-92.
- [30] G. Casella, and R. L. Berger. *Statistical inference* (Duxbury, Pacific Grove, CA, 2002), pp. 85-121.
- [31] S. M. Kay. *Fundamentals of Statistical Signal Processing: Estimation Theory* (Prentice Hall PTR, Upper Saddle River, NJ, 1993), pp. 500-539.